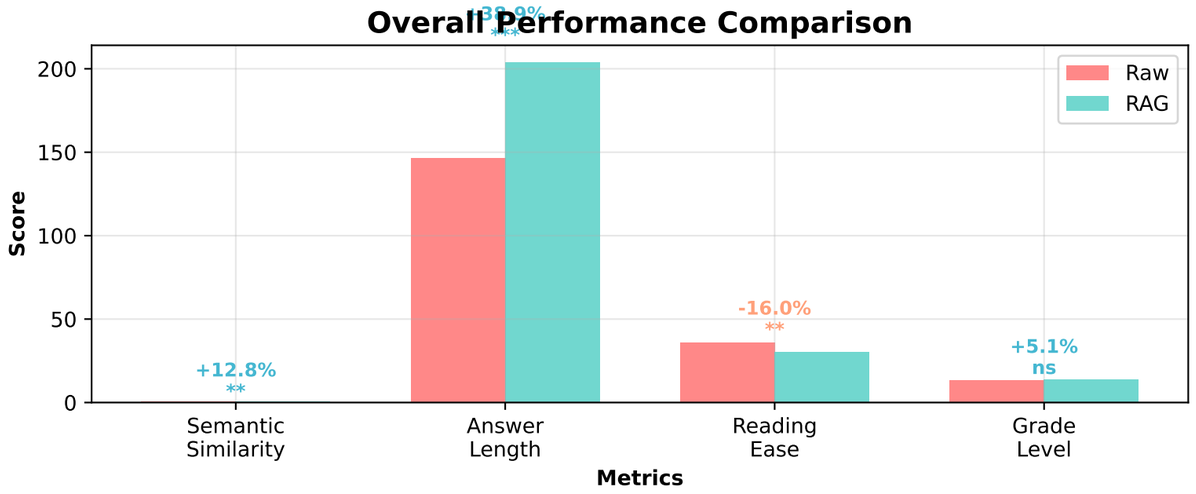
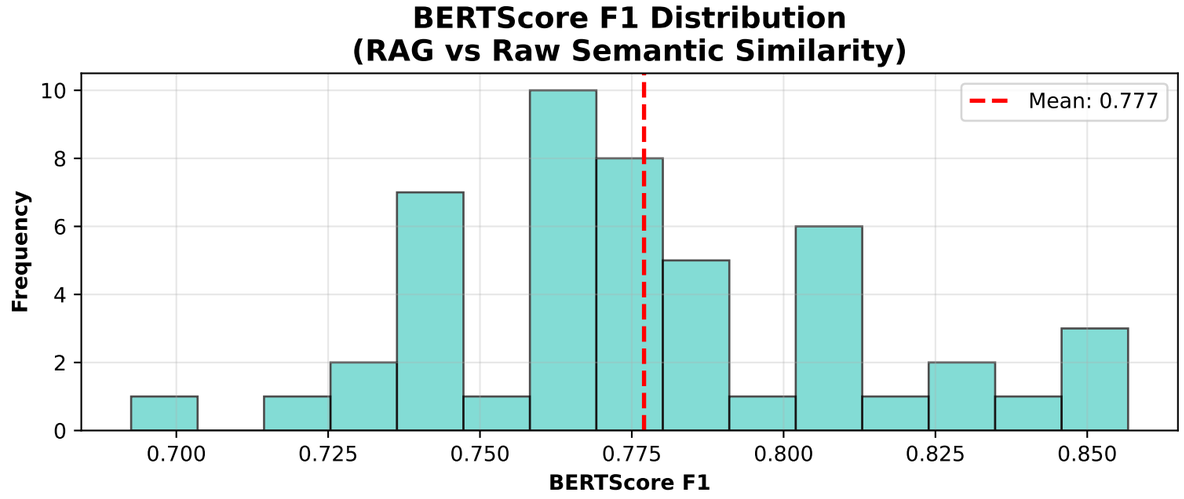
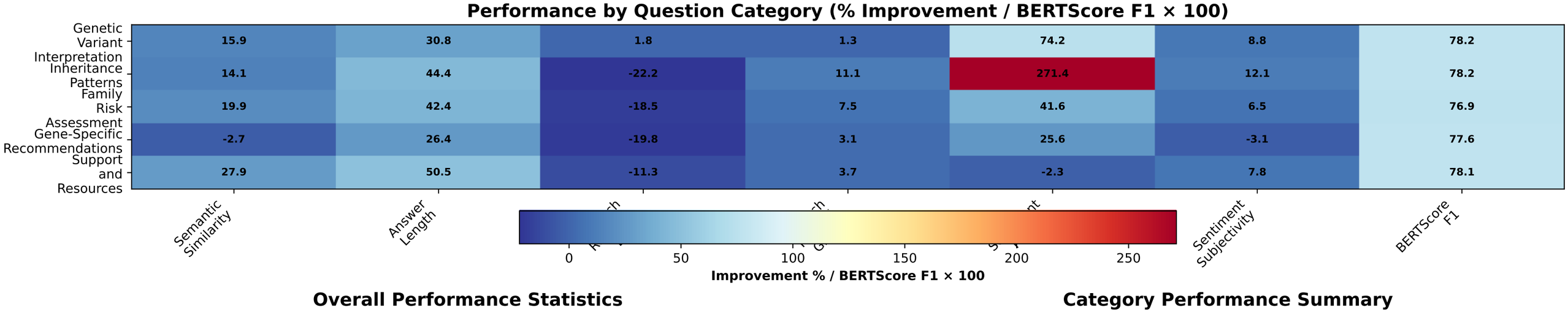
CGT Chatbot Performance Analysis: RAG vs Raw Approach
Comprehensive Results Dashboard







Overall Performance Statistics

Metric	Raw Mean	RAG Mean	Improvement %	P-value	Significant
Semantic Similarit	0.521	0.588	+12.8%	0.0043	Yes
Answer Length	146.735	203.878	+38.9%	0.0000	Yes
Flesch Reading Eas	35.898	30.161	-16.0%	0.0050	Yes
Flesch Kincaid Grad	13.376	14.053	+5.1%	0.0722	No
BERTScore F1	N/A	0.777	N/A	N/A	49/49

Category Performance Summary

Question Category	N Questions	Semantic Sim %	BERTScore F1	Significant Metrics
Genetic Variant...	5	+15.9%	0.782	1/6
Inheritance Pat...	7	+14.1%	0.782	2/6
Family Risk Ass...	11	+19.9%	0.769	2/6
Gene-Specific R...	14	-2.7%	0.776	1/6
Support and Res...	12	+27.9%	0.781	2/6

Raw Chatbot (Baseline) | RAG Chatbot (Enhanced) | Significant Improvement | Significant Decline